Econometrics

Basic Regression analysis with Time Series Data

Anna Donina

Lecture 9

The nature of time series data

- Temporal ordering of observations; may not be arbitrarily reordered
- Typical features: serial correlation/nonindependence of observations
- How should we think about the randomness in time series data?
  - The outcome of economic variables (e.g. GNP, Dow Jones) is uncertain; they should therefore be modeled as random variables
  - Time series are sequences of r.v. (= stochastic processes)
  - Randomness does not come from sampling from a population
  - „Sample“ = the one realized path of the time series out of the many possible paths the stochastic process could have taken

The nature of time series data

- **Example: US inflation and unemployment rates 1948-2003**

TABLE 10.1 Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–2003

Year	Inflation	Unemployment
1948	8.1	3.8
1949	−1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
.	.	.
.	.	.
.	.	.
1998	1.6	4.5
1999	2.2	4.2
2000	3.4	4.0
2001	2.8	4.7
2002	1.6	5.8
2003	2.3	6.0

© Ce

Here, there are only two time series. There may be many more variables whose paths over time are observed simultaneously.

<u>Time series analysis focuses on modeling the dependency of a variable on its own past, and on the present and past values of other variables.</u>

Examples of Time Series Regression Models

- **Static models**
  - In static time series models, the current value of one variable is modeled as the result of the current values of explanatory variables

- **Examples for static models**

$$inf_t = \beta_0 + \beta_1 unem_t + u_t$$

There is a contemporaneous relationship between unemployment and inflation (= Phillips-Curve).

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t$$

The current murder rate is determined by the current conviction rate, unemployment rate, and fraction of young males in the population.

Examples of Time Series Regression Models

- **Finite distributed lag models**
  - In finite distributed lag models, the explanatory variables are allowed to influence the dependent variable with a time lag

- **Example for a finite distributed lag model**
  - The fertility rate may depend on the tax value of a child, but for biological and behavioral reasons, the effect may have a lag

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

Children born per 1,000 women in year t

Tax exemption in year t

Tax exemption in year t-1

Tax exemption in year t-2

Examples of Time Series Regression Models

- **Interpretation of the effects in finite distributed lag models**

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t$$

- **Effect of a past shock on the current value of the dep. variable**

$$\frac{\partial y_t}{\partial z_{t-s}} = \delta_s$$

Effect of a transitory shock:
If there is a one time shock in a past period, the dep. variable will change temporarily by the amount indicated by the coefficient of the corresponding lag.

$$\frac{\partial y_t}{\partial z_{t-q}} + \cdots + \frac{\partial y_t}{\partial z_t} = \delta_1 + \cdots + \delta_q$$

Effect of permanent shock:
If there is a permanent shock in a past period, i.e. the explanatory variable permanently increases by one unit, the effect on the dep. variable will be the cumulated effect of all relevant lags. This is a long-run effect on the dependent variable.

Finite sample properties of OLS under classical assumptions

- **Assumption TS.1 (Linear in parameters)**

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t$$

The time series involved obey a linear relationship. The stochastic processes y_t, $x_{t1}, \ldots, x_{tk}$ are observed, the error process u_t is unobserved. The definition of the explanatory variables is general, e.g. they may be lags or functions of other explanatory variables.

- **Assumption TS.2 (No perfect collinearity)**

„In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others."

Finite sample properties of OLS under classical assumptions

- **Notation**

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tk} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

This matrix collects all the information on the complete time paths of all explanatory variables

The values of all explanatory variables in period number t

- **Assumption TS.3 (Zero conditional mean)**

$$E(u_t|\mathbf{X}) = 0$$

The mean value of the unobserved factors is unrelated to the values of the explanatory variables <u>in all periods</u>

Finite sample properties of OLS under classical assumptions

- **Discussion of assumption TS.3**

Exogeneity: $E(u_t|\mathbf{x}_t) = 0$ ← The mean of the error term is unrelated to the explanatory variables of the same period

Strict exogeneity: $E(u_t|\mathbf{X}) = 0$ ← The mean of the error term is unrelated to the values of the explanatory variables of all periods

- **Strict exogeneity is stronger than contemporaneous exogeneity**
  - TS.3 rules out feedback from the dep. variable on future values of the explanatory variables; this is often questionable esp. if explanatory variables „adjust“ to past changes in the dependent variable (*example: murder rate and police pc*)
  - If the error term is related to past values of the explanatory variables, one should include these values as contemporaneous regressors

Finite sample properties of OLS under classical assumptions

- **Theorem 10.1 (Unbiasedness of OLS)**

$$TS.1\text{–}TS.3 \quad \Rightarrow \quad E(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, \ldots, k$$

- **Assumption TS.4 (Homoscedasticity)**

$$Var(u_t|\mathbf{X}) = Var(u_t) = \sigma^2$$

← The volatility of the errors must not be related to the explanatory variables in any of the periods

  - A sufficient condition is that the volatility of the error is independent of the explanatory variables and that it is constant over time
  - In the time series context, homoscedasticity may also be easily violated, e.g. if the volatility of the dep. variable depends on regime changes (*example: T-bill rate and infl, deficit*)

Finite sample properties of OLS under classical assumptions

- **Assumption TS.5 (No serial correlation)**

$$Corr(u_t, u_s|\mathbf{X}) = 0, \; t \neq s$$

← Conditional on the explanatory variables, the unobserved factors must not be correlated over time

- **Discussion of assumption TS.5**
  - The assumption may easily be violated if, conditional on knowing the values of the indep. variables, omitted factors are correlated over time
  - Why was such an assumption not made in the cross-sectional case?
  - The assumption may also serve as substitute for the random sampling assumption if sampling a cross-section is not done completely randomly
  - In this case, given the values of the explanatory variables, errors have to be uncorrelated across cross-sectional units (e.g. states)

Finite sample properties of OLS under classical assumptions

- **Theorem 10.2 (OLS sampling variances)**

The same formula as in the cross-sectional case

Under assumptions TS.1 – TS.5:

$$Var(\hat{\beta}_j|\mathbf{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \quad j = 1, \ldots, k$$

- **Theorem 10.3 (Unbiased estimation of the error variance)**

$$TS.1 - TS.5 \quad \Rightarrow \quad E(\hat{\sigma}^2) = \sigma^2$$

Finite sample properties of OLS under classical assumptions

- **Theorem 10.4 (Gauss-Markov Theorem)**
  - Under assumptions TS.1 – TS.5, the OLS estimators have the minimal variance of all linear unbiased estimators of the regression coefficients
  - This holds conditional as well as unconditional on the regressors

- **Assumption TS.6 (Normality)**

This assumption implies TS.3 – TS.5

$$u_t \sim N(0, \sigma^2) \quad \text{independently of x}$$

- **Theorem 10.5 (Normal sampling distributions)**
  - Under assumptions TS.1 – TS.6, the OLS estimators have the usual normal distribution (conditional on x).

  The usual F- and t-tests are valid.

Finite sample properties of OLS under classical assumptions

- **Example: Static Phillips curve**

$$\widehat{inf}_t = \underset{(1.72)}{1.42} + \underset{(.289)}{.468}\ unem_t$$

$$n = 49, R^2 = .053, \bar{R}^2 = .033$$

Contrary to theory, the estimated Phillips Curve does not suggest a tradeoff between inflation and unemployment

- **Discussion of CLM assumptions**

TS.1: $inf_t = \beta_0 + \beta_1 unem_t + u_t$

The error term contains factors such as monetary shocks, income/demand shocks, oil price shocks, supply shocks, or exchange rate shocks

TS.2: A linear relationship might be restrictive, but it should be a good approximation. Perfect collinearity is not a problem as long as unemployment varies over time.

Finite sample properties of OLS under classical assumptions

- **Discussion of CLM assumptions (cont.)**

TS.3: $E(u_t|unem_1, \ldots, unem_n) = 0$ ← Easily violated

$unem_{t-1} \uparrow \rightarrow \; u_t \downarrow$ ← For example, past unemployment shocks may lead to future demand shocks which may dampen inflation

$u_{t-1} \uparrow \rightarrow \; unem_t \uparrow$ ← For example, an oil price shock means more inflation and may lead to future increases in unemployment

TS.4: $Var(u_t|unem_1, \ldots, unem_n) = \sigma^2$ ← Assumption is violated if monetary policy is more „nervous“ in times of high unemployment

TS.5: $Corr(u_t, u_s|unem_1, \ldots, unem_n) = 0$ ← Assumption is violated if exchange rate influences persist over time (they cannot be explained by unemployment)

TS.6: $u_t \sim N(0, \sigma^2)$ ← Questionable

Finite sample properties of OLS under classical assumptions

- **Example: Effects of inflation and deficits on interest rates**

Interest rate on 3-months T-bill

Government deficit as percentage of GDP

$$\widehat{i3}_t = \underset{(0.43)}{1.73} + \underset{(.082)}{.606}\ inf_t + \underset{(.118)}{.513}\ def_t$$

$$n = 56, R^2 = .602, \bar{R}^2 = .587$$

- **Discussion of CLM assumptions**

TS.1: $i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t$

The error term represents other factors that determine interest rates in general, e.g. business cycle effects

TS.2: A linear relationship might be restrictive, but it should be a good approximation. Perfect collinearity will seldomly be a problem in practice.

Finite sample properties of OLS under classical assumptions

- **Discussion of CLM assumptions (cont.)**

TS.3: $E(u_t|inf_1, \ldots, inf_n, def_1, \ldots, def_n) = 0$ ← Easily violated

$def_{t-1} \uparrow \rightarrow \; u_t \uparrow$ ← For example, past deficit spending may boost economic activity, which in turn may lead to general interest rate rises

$u_{t-1} \uparrow \rightarrow \; inf_t \uparrow$ ← For example, unobserved demand shocks may increase interest rates and lead to higher inflation in future periods

TS.4: $Var(u_t|inf_1, \ldots, def_n) = \sigma^2$ ← Assumption is violated if higher deficits lead to more uncertainty about state finances and possibly more abrupt rate changes

TS.5: $Corr(u_t, u_s|inf_1, \ldots, def_n) = 0$ ← Assumption is violated if business cylce effects persist across years (and they cannot be completely accounted for by inflation and the evolution of deficits)

TS.6: $u_t \sim N(0, \sigma^2)$ ← Questionable

Using Dummy Explanatory Variables in Time Series

Children born per 1,000 women in year t

Tax exemption in year t

Dummy for World War II years (1941-45)

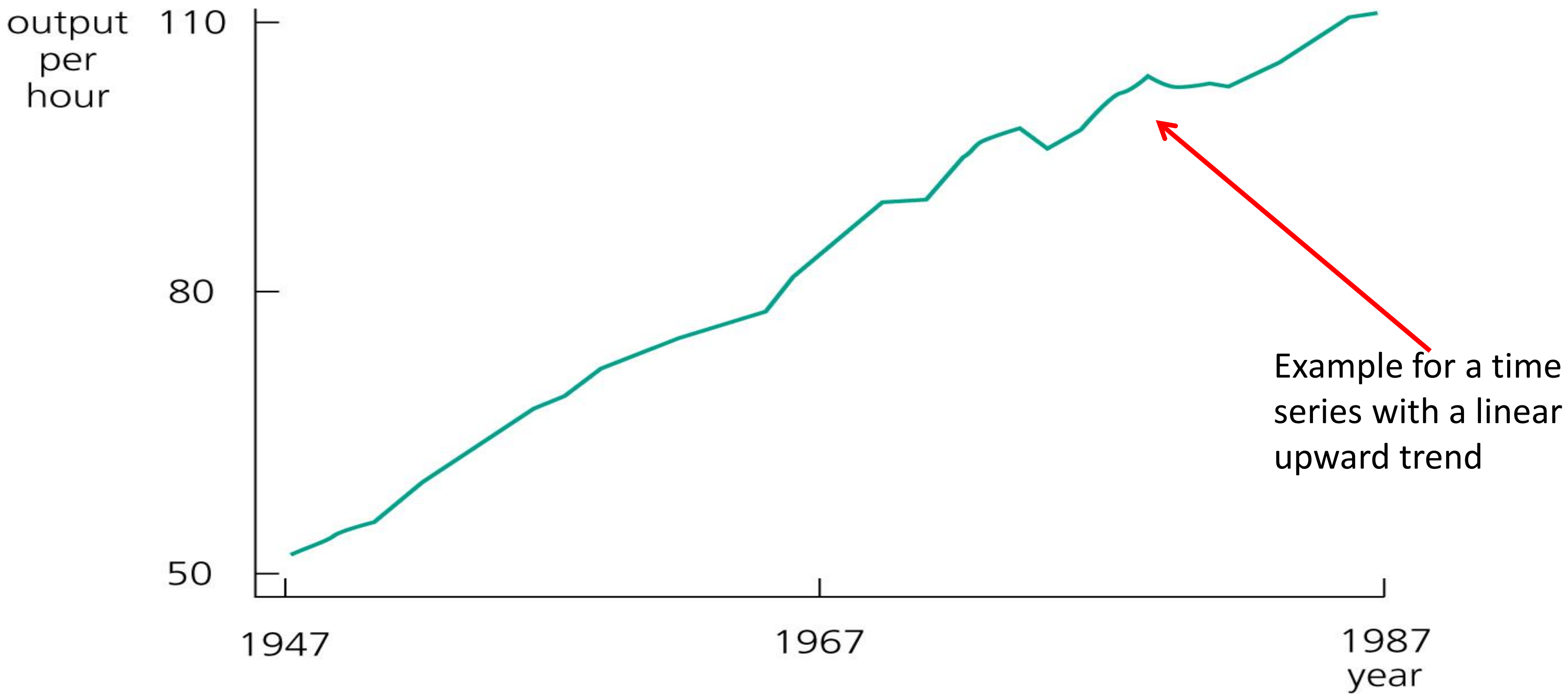
Dummy for availabity of contraceptive pill (1963-present)

$$\widehat{gfr}_t = \underset{(3.68)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24}\, ww2_t - \underset{(4.08)}{31.59}\, pill_t$$

$$n = 72, R^2 = .473, \bar{R}^2 = .450$$

- **Interpretation**
  - During World War II, the fertility rate was temporarily lower
  - It has been permanently lower since the introduction of the pill in 1963

Time series with trends



Time series with trends

- **Modelling a linear time trend**

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta y_t) = E(y_t - y_{t-1}) = \alpha_1$$

$\partial y_t / \partial t = \alpha_1$ ← Abstracting from random deviations, the dependent variable increases by a constant amount per time unit

$E(y_t) = \alpha_0 + \alpha_1 t$ ← Alternatively, the expected value of the dependent variable is a linear function of time

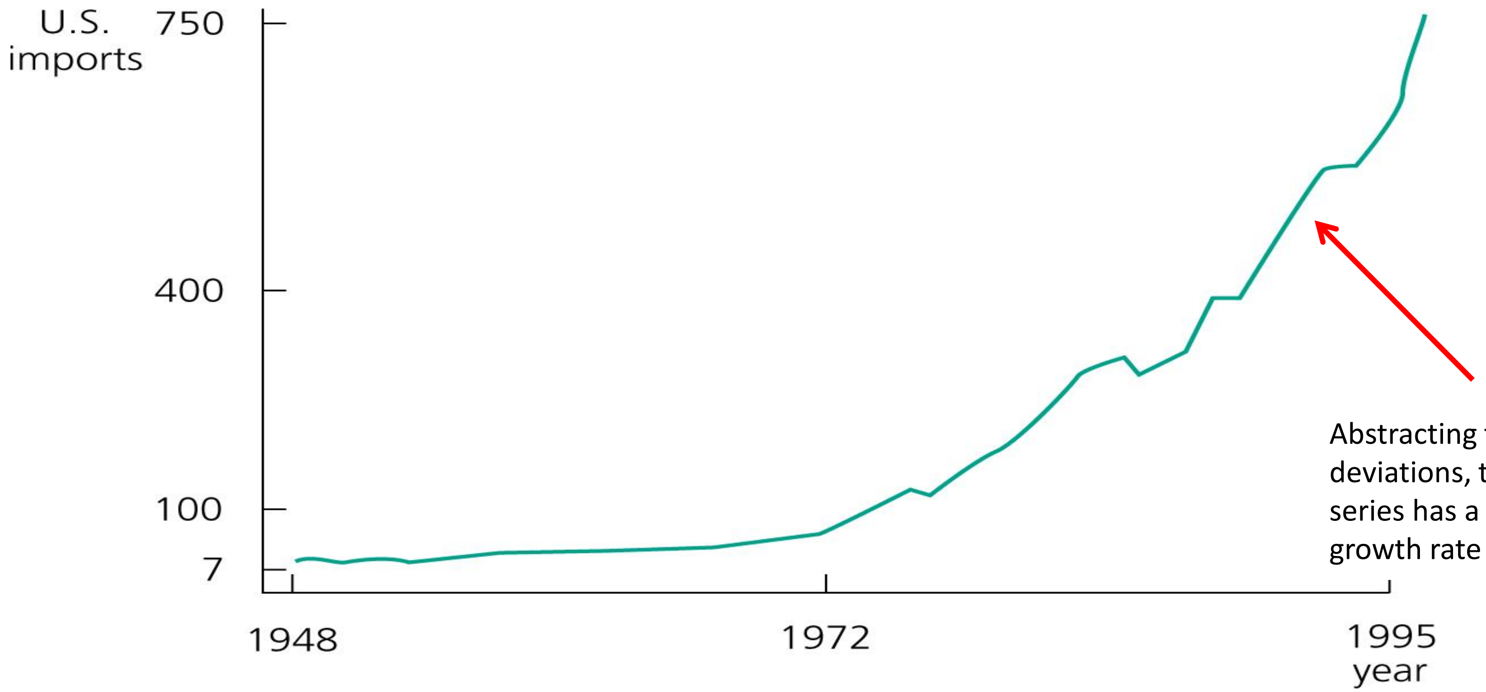
- **Modelling an exponential time trend**

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta \log(y_t)) = \alpha_1$$

$(\partial y_t / y_t) / \partial t = \alpha_1$ ← Abstracting from random deviations, the dependent variable increases by <u>a constant percentage</u> per time unit

Time series with trends

- **Example for a time series with an exponential trend**



Abstracting from random deviations, the time series has a constant growth rate

Time series with trends

- **Using trending variables in regression analysis**
  - If trending variables are regressed on each other, a spurious relationship may arise if the variables are driven by a common trend
  - In this case, it is important to include a trend in the regression
- **Example: Housing investment and prices**

Per capita housing investment

Housing price index

$$\widehat{\log}(invpc) = -\underset{(.043)}{.550} + \underset{(.382)}{1.241}\log(price)$$

$$n = 42, R^2 = .208, \bar{R}^2 = .189$$

It looks as if investment and prices are positively related

Time series with trends

- **Example: Housing investment and prices (cont.)**

$$\widehat{\log}(invpc) = \underset{(.136)}{-\ .913} + \underset{(.679)}{.381}\ \log(price) + \underset{(.0035)}{.0098}\ t$$

$$n = 42, R^2 = .341, \bar{R}^2 = .307$$

There is no significant relationship between price and investment anymore

- **When should a trend be included?**
  - If the dependent variable displays an obvious trending behaviour
  - If both the dependent and some independent variables have trends
  - <u>If only some of the independent variables have trends</u>; their effect on the dep. var. may only be visible after a trend has been substracted

Time series with trends

- **A Detrending interpretation of regressions with a time trend**
  - It turns out that the OLS coefficients in a regression including a trend are the same as the coefficients in a regression without a trend but where all the variables have been detrended before the regression
  - This follows from the general interpretation of multiple regressions
- **Computing R-squared when the dependent variable is trending**
  - Due to the trend, the variance of the dep. var. will be overstated
  - It is better to first detrend the dep. var. and then run the regression on all the indep. variables (plus a trend if they are trending as well)
  - The R-squared of this regression is a more adequate measure of fit

Time series with trends

- <u>**Modelling seasonality in time series**</u>
- **A simple method is to include a set of seasonal dummies:**

$$y_t = \beta_0 + \delta_1 feb_t + \delta_2 mar_t + \delta_3 apr_t + \cdots + \delta_{11} dec_t$$
$$+ \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$$

=1 if obs. from december
=0 otherwise

- **Similar remarks apply as in the case of deterministic time trends**
  - The regression coefficients on the explanatory variables can be seen as the result of first deseasonalizing the dependent and the explanatory variables
  - An R-squared that is based on first deseasonalizing the dependent variable may better reflect the explanatory power of the explanatory variables